Exercice 1 Dégraissage 4 points

Une entreprise emploie 200 ouvriers dont le salaire annuel est de 15 000 euros et deux cadres dont le salaire annuel est de 45 000 euros. A la suite d'une restructuration de l'entreprise et de nombreux licenciements, le responsable du personnel peut annoncer aux syndicats que les salaires ont progresser de plus de 15 % en moyenne et dans le même temps pour rassurer les actionnaires de l'entreprise en faisant état d'une augmentation limitée à 5% pour les cadres et 10% pour les ouvriers.

1. Comment cela est-il possible ?

Exercice 2 Chose promise... 3 points

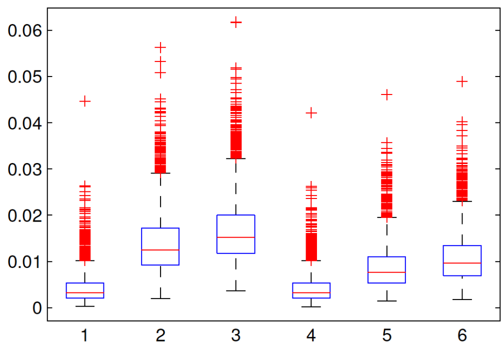
voici le début d'une session Matlab :

```
[n,m] = size(v);
          n = 100
          m = 1
[n,m] = size(u);
         n = 100
          m = 1
[n,m] = size(x);
          n = 100
           m = 1
```

1. pour chacune des instructions suivante indiquez :
   a) la signification de l'instruction (en une phrase),
   b) la nature du résultat et sa taille.

```
M = [ones(n,1) u v];
[n,p] = size(M);
y = (M'*M)\(M'*x);
a = M*y
e = x-a
H = M*inv(M'*M)*M'
H*H
sum(e)
```

Exercice 3 Bigote 3 points



La figure ci-dessus présente pour 6 méthodes différentes (abscisse) les performances en terme de probabilité d'erreur sur des jeux de test (ordonnée). Plus la probabilité d'erreur est faible, meilleure est la méthode.

1. quels commentaires vous inspire cette figure ? quelle(s) est (sont) les meilleure(s) méthode(s) ? quelles sont celles qui sont équivalentes et pourquoi ?

Exercice 4 Retour de l'épée fatiguée 10 points

À la suite de l'énoncé, vous trouverez la copie d'une session Matlab. On y trouve des lignes de code qui commencent par le chargement d'un tableau de données X comportant $p = 4$ variables et $n = 25$ observations.

En annexe de cet énoncé, vous trouverez deux figures représentant le nuage points des deux premières variables du tableau X. Ces figures vous serviront à répondre aux deux premières questions de cet exercice.

Rappel : la fonction matlab `[U,d] = eig(M)` calcule les vecteurs propres (U) et les valeurs propres associées (sur la diagonale de d) de la matrice M.

1. représentez les enveloppes convexes successives et indiquez où se situe la médiane de Tukey du nuage de points,
2. représentez le sac médian du nuage de points,
3. y a-t'il des points aberrants ? quel impact pour la suite de l'analyse statistique ?
4. quelle(s) est/sont, selon vous, l'(es) étape(s) indispensable(s) à appliquer avant d'effectuer une ACP ?
5. quelle sont les moyennes et variances de X et de A ?
6. quelle relation permet de représenter les données par rapport à un axe de l'ACP ?
7. caractérisez la qualité de la représentation du nuage de points par rapport à la première composante de l'ACP ?
8. représentez les variables dans le plan des deux premières composantes principales. Quelle interprétation donnez vous à cette représentation ?

```
load X

X =

  -0.5410   -0.6331    0.3745    0.7852
   0.6451   -0.1519    0.9507    0.1997
   0.7272   -1.2195    0.7320    0.5142
   0.3686   -2.1837    0.5987    0.5924
  -1.5703    0.9193    0.1560    0.0465
  -2.1499    0.3417    0.1560    0.6075
  -0.0876   -2.6452    0.0581    0.1705
  -2.0000    1.5000    0.8662    0.0651
  -2.6466   -1.4144    0.6011    0.9489
   1.1217   -2.5342    0.7081    0.9656
   1.8794    1.0015    0.0206    0.8084
  -0.1118    2.3506    0.9699    0.3046
  -3.2037   -2.0683    0.8324    0.0977
  -2.8480    1.5765    0.2123    0.6842
   1.6140   -3.1144    0.1818    0.4402
  -2.1306   -3.5424    0.1834    0.1220
   3.0000   -1.0000    0.3042    0.4952
   2.0483   -3.5186    0.5248    0.0344
  -1.3114   -4.6953    0.4319    0.9093
  -3.6899   -3.7823    0.2912    0.2588
  -3.0942   -4.3754    0.6119    0.6625
  -1.5000    4.0000    0.1395    0.3117
  -2.9431    3.4100    0.2921    0.5201
  -3.9392   -4.3146    0.3664    0.5467
   6.3000   -8.6000    0.4561    0.1849

mean(X) = -0.6425   -1.3878    0.4408    0.4510
[n p] = size(X);
```

```
cov(X) = 5.9875   -2.1368    0.0211   -0.0569
        -2.1368    8.4077   -0.0333    0.0090
         0.0211   -0.0333    0.0813   -0.0071
        -0.0569    0.0090   -0.0071    0.0890

A = (X - ones(n,1)*mean(X))./(ones(n,1)*std(X));
cov(A) = 1.0000   -0.3012    0.0302   -0.0779
        -0.3012    1.0000   -0.0403    0.0104
         0.0302   -0.0403    1.0000   -0.0832
        -0.0779    0.0104   -0.0832    1.0000

[U,d]=eig(X'*X)
U =-0.0076    0.0937    0.9617   -0.2573
    0.0027    0.1026    0.2478    0.9634
    0.7205    0.6870   -0.0761   -0.0556
   -0.6934    0.7133   -0.0887   -0.0512

diag(d) =   2.1989   8.4937   147.8644  259.4253

[Vp,valp]=eig(A'*A)
Vp =-0.7035   -0.1053    0.1665   -0.6828
    -0.6864   -0.1321   -0.2734    0.6608
    -0.0668    0.7115   -0.6691   -0.2041
    -0.1718    0.6821    0.6707    0.2354

diag(valp) = 16.5600 22.1583 25.4257 31.8561
```

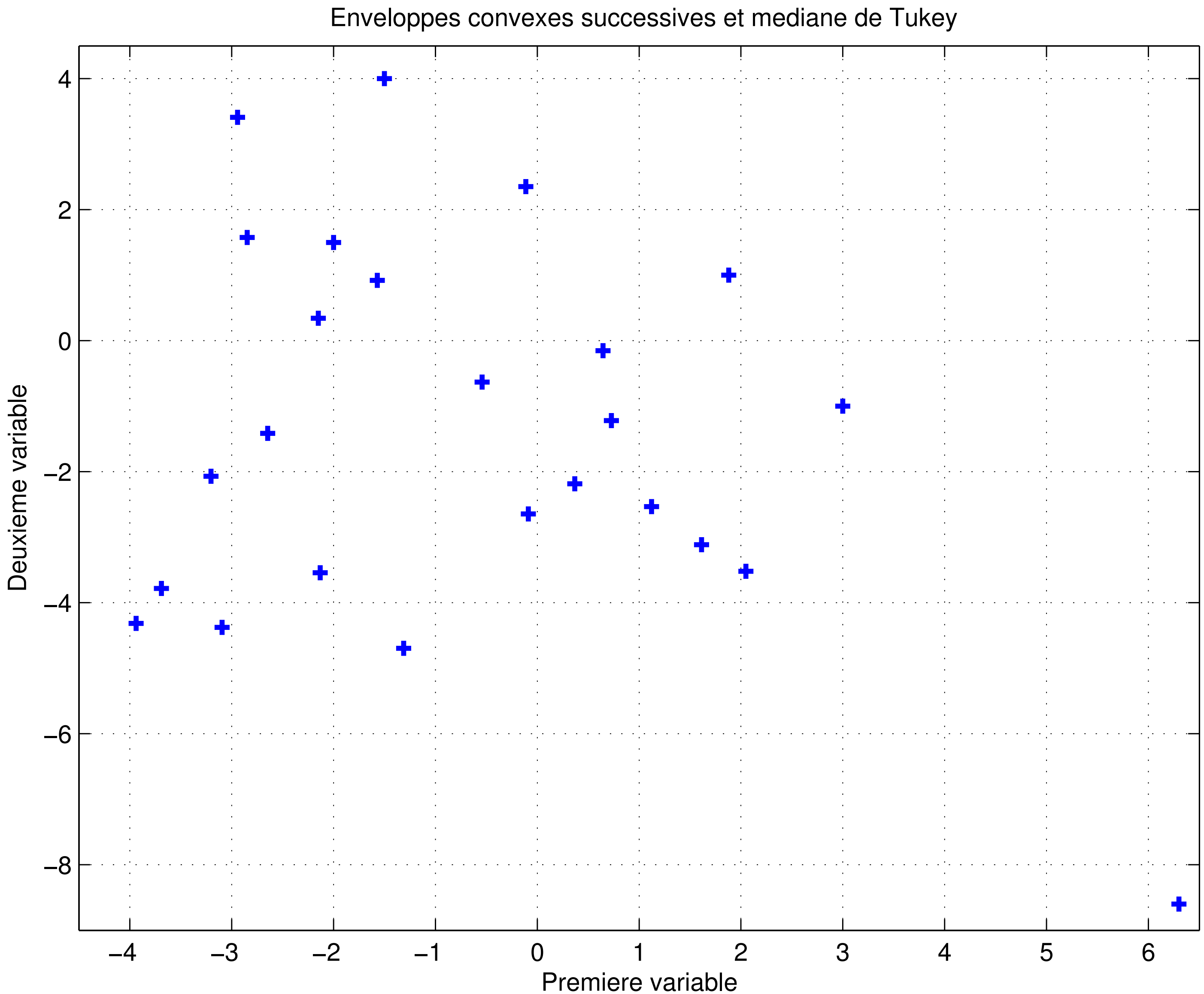


FIGURE 1 – Représentez graphiquement la médiane de Tukey et les enveloppes convexes successives.

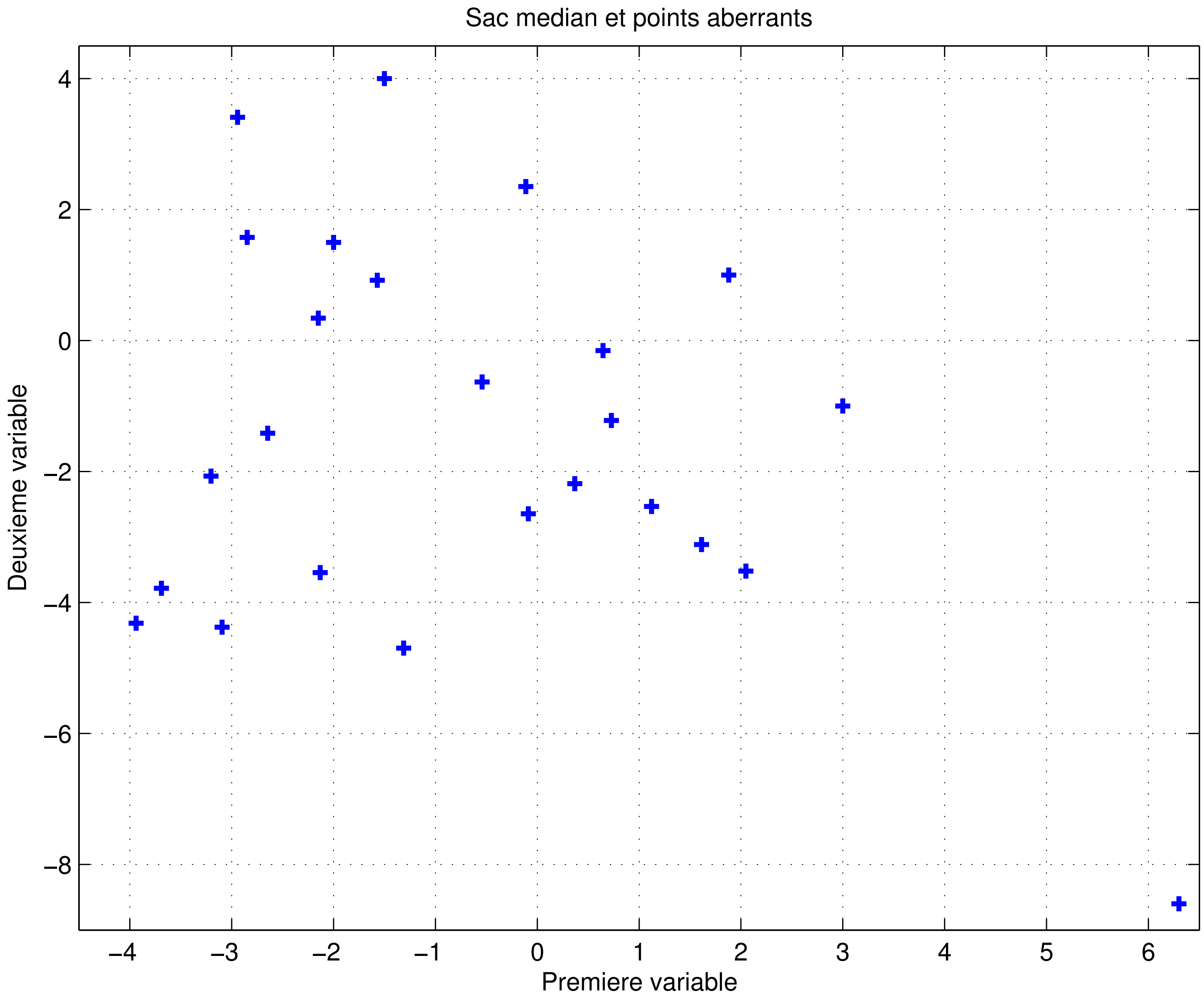


FIGURE 2 – Tracez le sace médian et repérez les éventuelles points aberrants sur cette figure.